Unsupervised State Representation Learning in Atari

Ankesh Anand*, Evan Racah*, Sherjil Ozair*, Yoshua Bengio, Marc-Alexandre Cote, R Devon Hjelm

*Equal Contribution. {anandank, racaheva, ozairs}@mila.quebec

Overview

State representation learning, or the ability to capture **latent generative factors** of an environment, is crucial for building intelligent agents that can perform a wide variety of tasks.

★ We introduce **SpatioTemporal DeepInfoMax (ST-DIM)** which maximizes predictive mutual-information to learn high-level concepts in a scene:
- ✓ without labels or rewards;
- ✓ without modelling pixels directly.

★ We also introduce the **Atari Annotated RAM Interface (AtariARI)**, which exposes the ground truth semantic information present in the RAM state. We use AtariARI to **evaluate** state representations based on how well they capture the ground truth state variables.

Atari Annotated Ram Interface

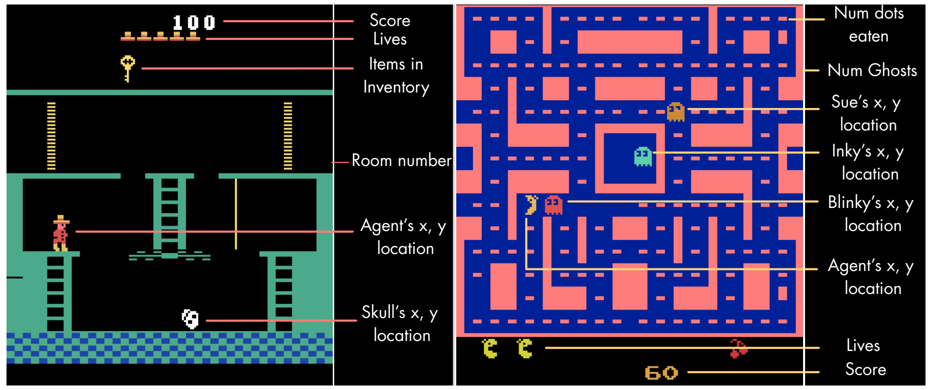


Figure 1: Identified RAM states for Montezuma's Revenge (left) and MsPacman (right).

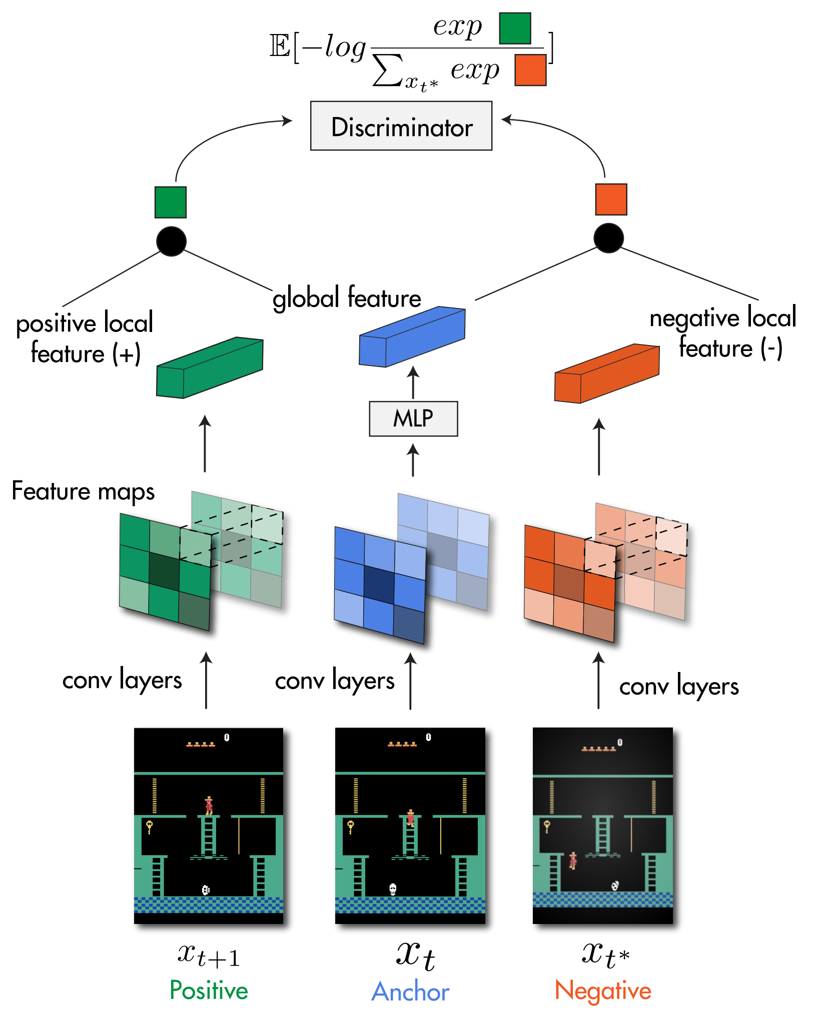
★ We identify important state variables from source code of 22 games. State variables include location of player, location of items of interest (keys, doors, ..), enemies, score etc.

★ Representations are evaluated using **linear probing**, i.e. the accuracy of linear classifiers trained to predict each latent generative factor from the learned representations.

★ Gradients are **not backpropagated** through the encoder network during evaluation

Spatio-Temporal Deep Infomax

ST-DIM learns representations by maximizing mutual information of representations across spatial and temporal axes.



Insights

★ Contrastive methods (ST-DIM and CPC) perform better than generative methods (VAE and PIXEL-PRED) across all semantic categories. Overall, ST-DIM largely outperforms other methods in terms of probe F1 score.

★ ST-DIM excels at capturing small objects, and is able to learn multiple generative factors by leveraging local losses.

★ The AtariARI benchmark can be used to systematically evaluate different state representation learning techniques.

Results

We consider two different modes for collecting the data:

- random agent (actions are picked uniformly at random)
- pretrained PPO agent (NB: results are in the paper).;

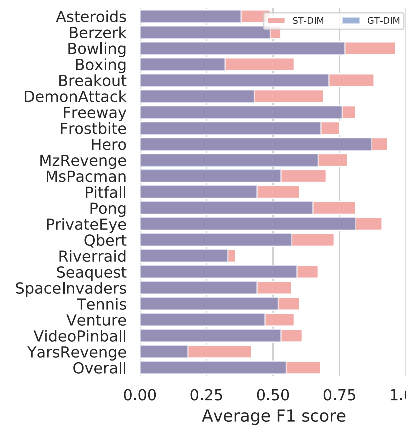
Table 1: Probe F1 scores for data collected by random agents

Game	maj-clf	random-cnn	vae	pixel-pred	cpc	st-dim	supervised
asteroids	0.28	0.34	0.36	0.34	0.42	**0.49**	0.52
berzerk	0.18	0.43	0.45	**0.55**	**0.56**	0.53	0.68
bowling	0.33	0.48	0.50	0.81	0.90	**0.96**	0.95
boxing	0.01	0.19	0.20	0.44	0.29	**0.58**	0.83
breakout	0.17	0.51	0.57	0.70	0.74	**0.88**	0.94
demonattack	0.16	0.26	0.25	0.32	0.57	**0.69**	0.83
freeway	0.01	0.50	0.26	**0.81**	0.47	**0.81**	0.98
frostbite	0.08	0.57	0.01	0.72	**0.76**	**0.75**	0.85
hero	0.22	0.75	0.51	0.74	0.90	**0.93**	0.98
montezuma	0.08	0.68	0.69	0.74	0.75	**0.78**	0.87
mspacman	0.10	0.48	0.38	**0.74**	0.65	0.70	0.87
pitfall	0.07	0.34	0.56	0.44	0.46	**0.60**	0.83
pong	0.10	0.17	0.09	0.70	0.71	**0.81**	0.87
privateeye	0.23	0.70	0.71	0.83	0.81	**0.91**	0.97
qbert	0.29	0.49	0.49	0.52	0.65	**0.73**	0.76
riverraid	0.04	0.34	0.26	**0.41**	**0.40**	0.36	0.57
seaquest	0.29	0.57	0.56	0.62	**0.66**	**0.67**	0.85
spaceinvaders	0.14	0.41	0.52	**0.57**	0.54	**0.57**	0.75
tennis	0.09	0.41	0.29	0.57	**0.60**	**0.60**	0.81
venture	0.09	0.36	0.38	0.46	0.51	**0.58**	0.68
videopinball	0.09	0.37	0.45	0.57	0.58	**0.61**	0.82
yarsrevenge	0.01	0.22	0.08	0.19	0.39	**0.42**	0.74
mean	0.14	0.44	0.39	0.58	0.60	**0.68**	0.83

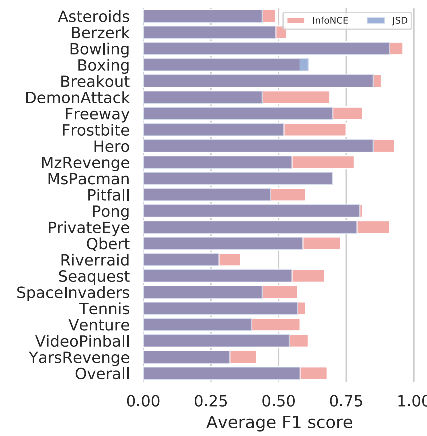
Average Probe F1 scores for data collected by random agents

Category	maj-clf	cnn	vae	pixel-pred	cpc	st-dim	supervised
Small Loc.	0.14	0.19	0.17	0.31	0.42	**0.51**	0.69
Agent Loc.	0.12	0.31	0.30	0.48	0.43	**0.58**	0.83
Other Loc.	0.14	0.50	0.36	0.61	0.66	**0.69**	0.81
Score/Clock/Lives	0.13	0.58	0.53	0.76	0.83	**0.86**	0.93
Misc.	0.26	0.59	0.65	0.70	0.71	**0.74**	0.86

Different Ablations for the ST-DIM model



Effect of Spatial Loss



InfoNCE vs JSD

Arxiv:

Github: